**Arctic Network**
**Inventory & Monitoring Program**

U.S. Department of the Interior

**Data Management**
**Standard Operating Procedure**
NPS/ARCN/DMSOP-2009-03

# Exploratory Data Analysis Using SQL Server Management Studio

***Overview of documenting and archiving datasets***

## Summary

Arctic Network uses SQL Server as an enterprise database for storage and retrieval of monitoring data. Where analysis is routine, many queries and analyses can be built into the SQL Server or any of a variety of front end applications. Exploratory data analysis, however, presents a challenge to users who are accustomed to being able to build ad-hoc queries on the fly using their preferred front end application, usually Microsoft Access. This SOP outlines the use of SQL Server Management Studio to perform ad-hoc querying of SQL Server databases.

## Contents



## Introduction

This document describes the use of Microsoft SQL Server Management Studio Express for ad-hoc querying and data analysis. This is a 'get-acquainted' tutorial. Readers are encouraged to use the software's documentation and internet resources for more advanced topics.

## Obtaining Microsoft SQL Server Management Studio

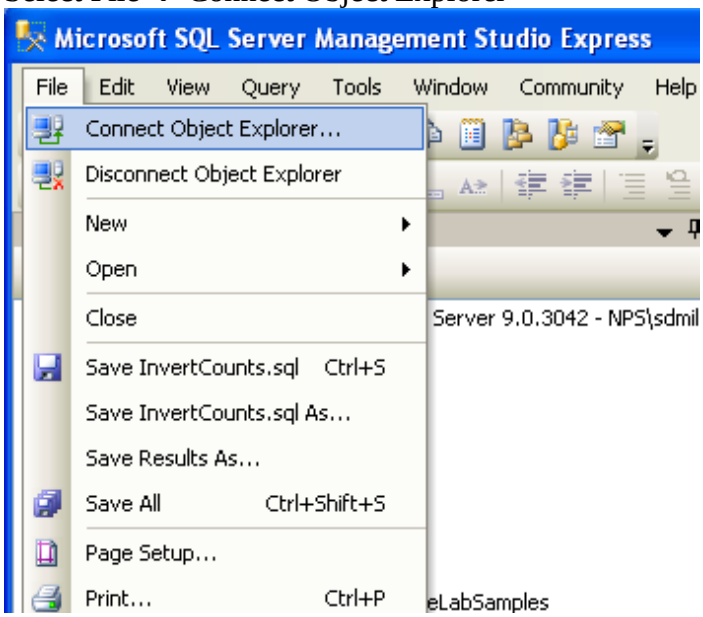
Download Microsoft SQL Server Management Studio Express at no charge from the Microsoft website

(http://www.microsoft.com/downloads/details.aspx?FamilyId=C243A5AE-4BD1-4E3D-94B8-5A0F62BF7796&displaylang=en).

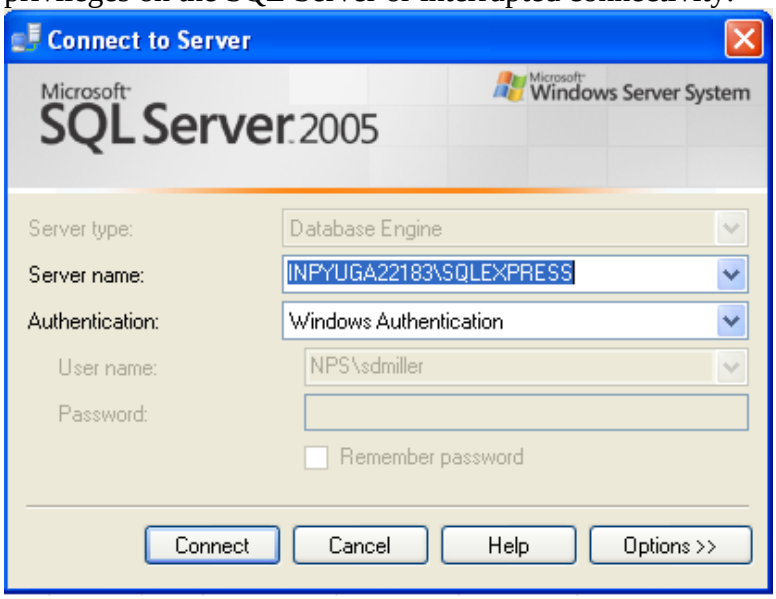
Install Microsoft SQL Server Management Studio
Open the program.

## Connecting to the Arctic Network SQL Server

Select File → Connect Object Explorer

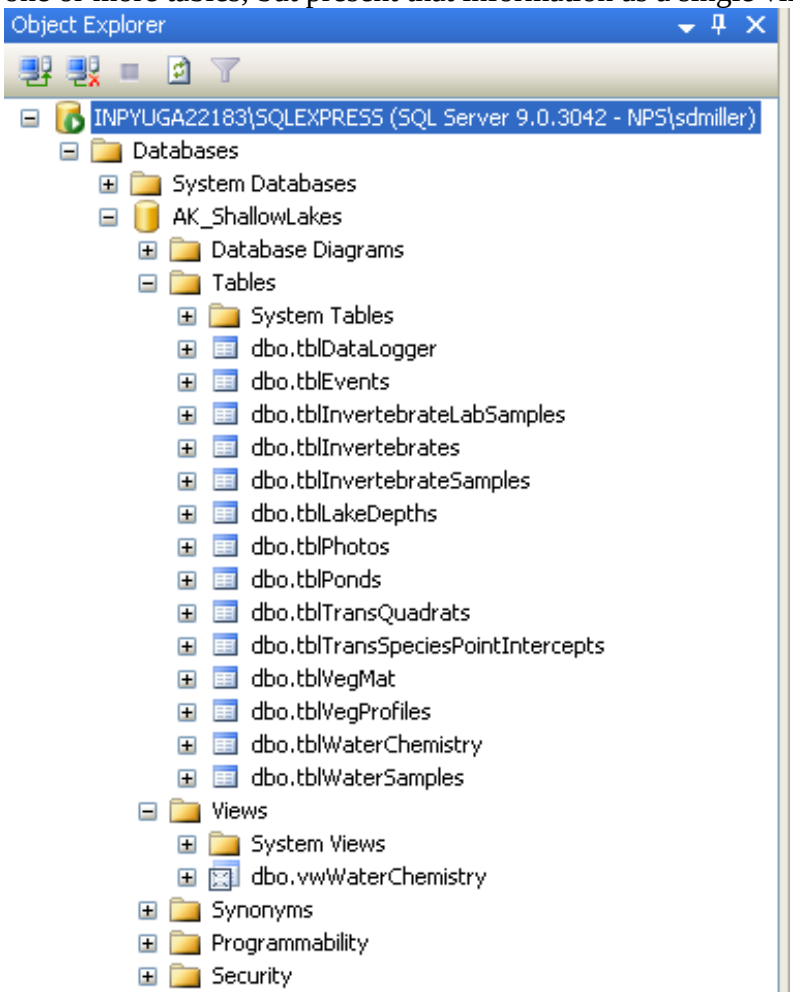



Input the name of the SQL Server, select "Windows Authentication" if it is not already selected and press connect. Connection errors are most likely caused by insufficient privileges on the SQL Server or interrupted connectivity.



# Viewing Tables and Views (Queries)

The database objects of most interest to users are tables and views. Tables contain the both the structure and values of the data. Views are very similar to saved queries in a program such as Microsoft Access. Views are precompiled queries that pull data from one or more tables, but present that information as a single virtual table.`




You may view table and view contents by right-clicking on them and selecting the appropriate option from the context menu.

# Querying Data

Querying data in Microsoft SQL Server Management Studio is similar to what you would do in Microsoft Access once you gain some familiarity with a few tools. The two major tools for querying are queries and views.

## *Should I make a query or a view?*

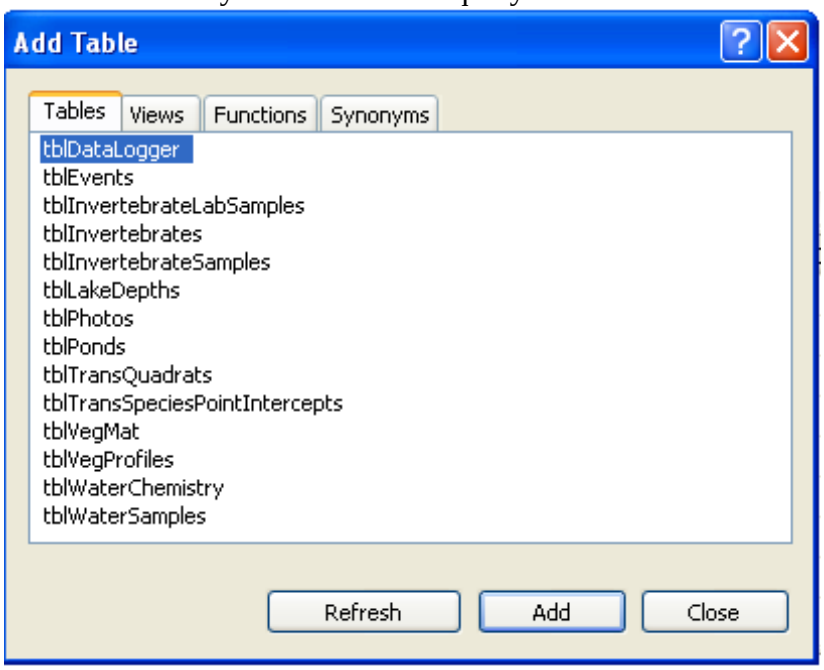
Queries and views are basically the same with one difference; a view, if saved, becomes a part of the database, a query does not. For exploratory data analysis one might think a query to be better than a view, but this is not the case for the simple reason that the view editor has a much better interface for ad-hoc data viewing. **I recommend a view for**

**exploratory data analysis.** If you are sure you will never use the view again you may simply delete it or choose not to save it.

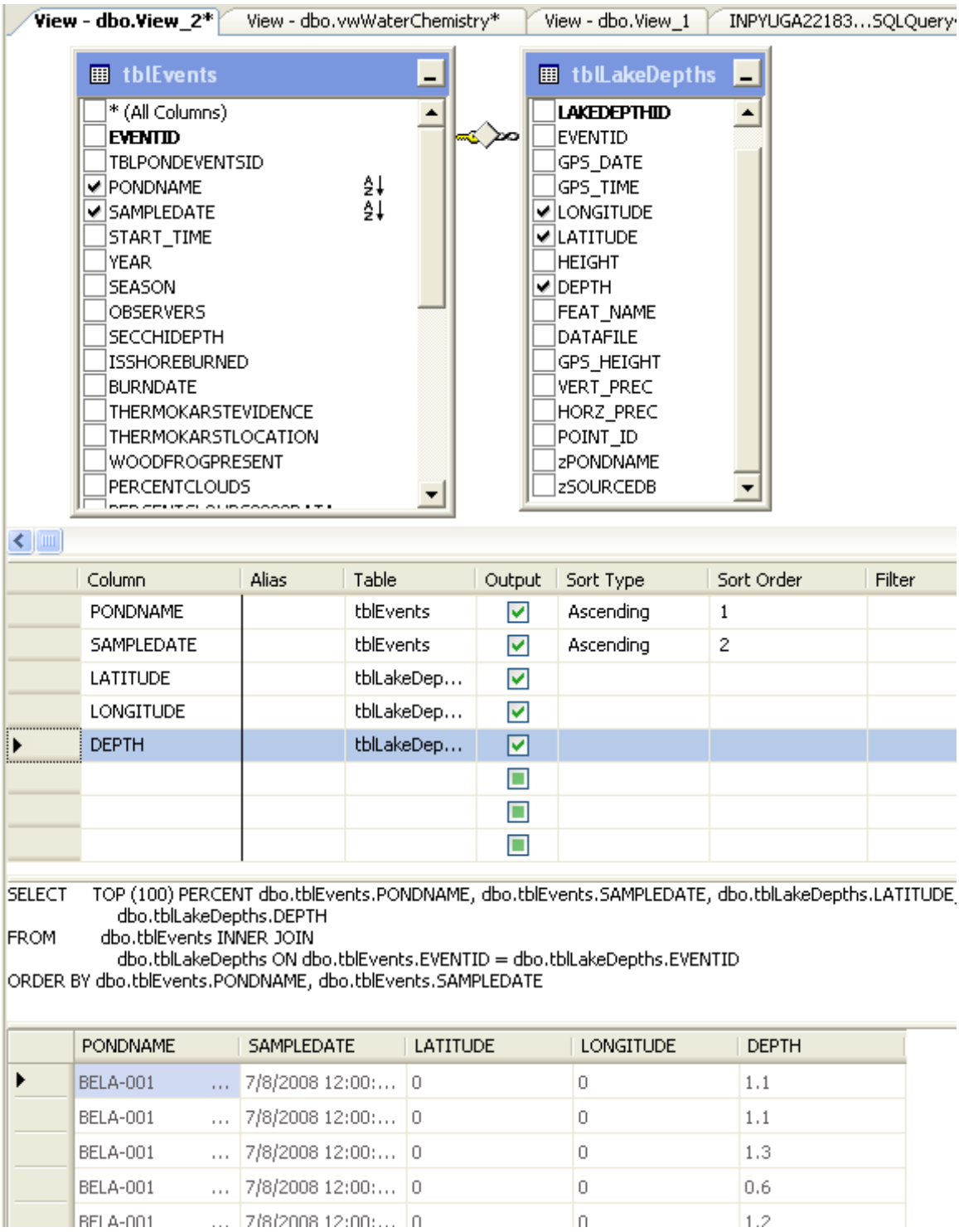
### *How to create a view*

Right click on the 'Views' node under your database.


Select the tables you would like to query



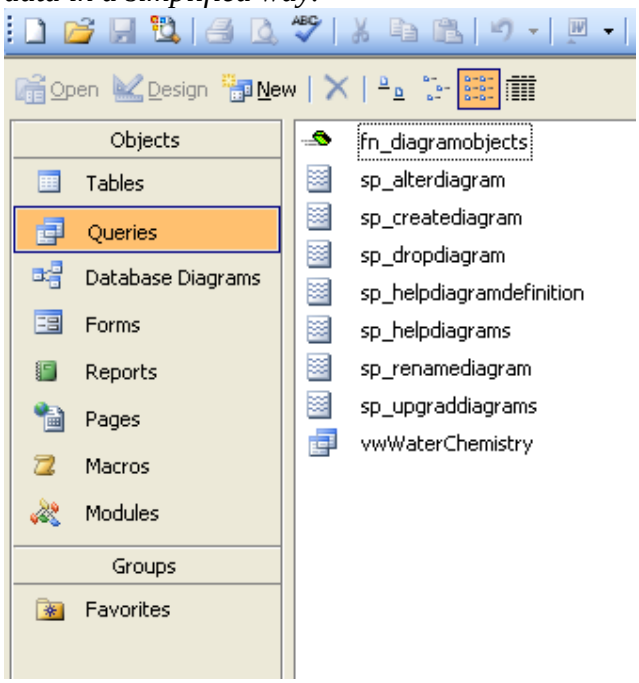

The View Editor Appears. The view editor allows you to see your tables and relationships, select, order and filter columns and view the results.




## *Saving a view*

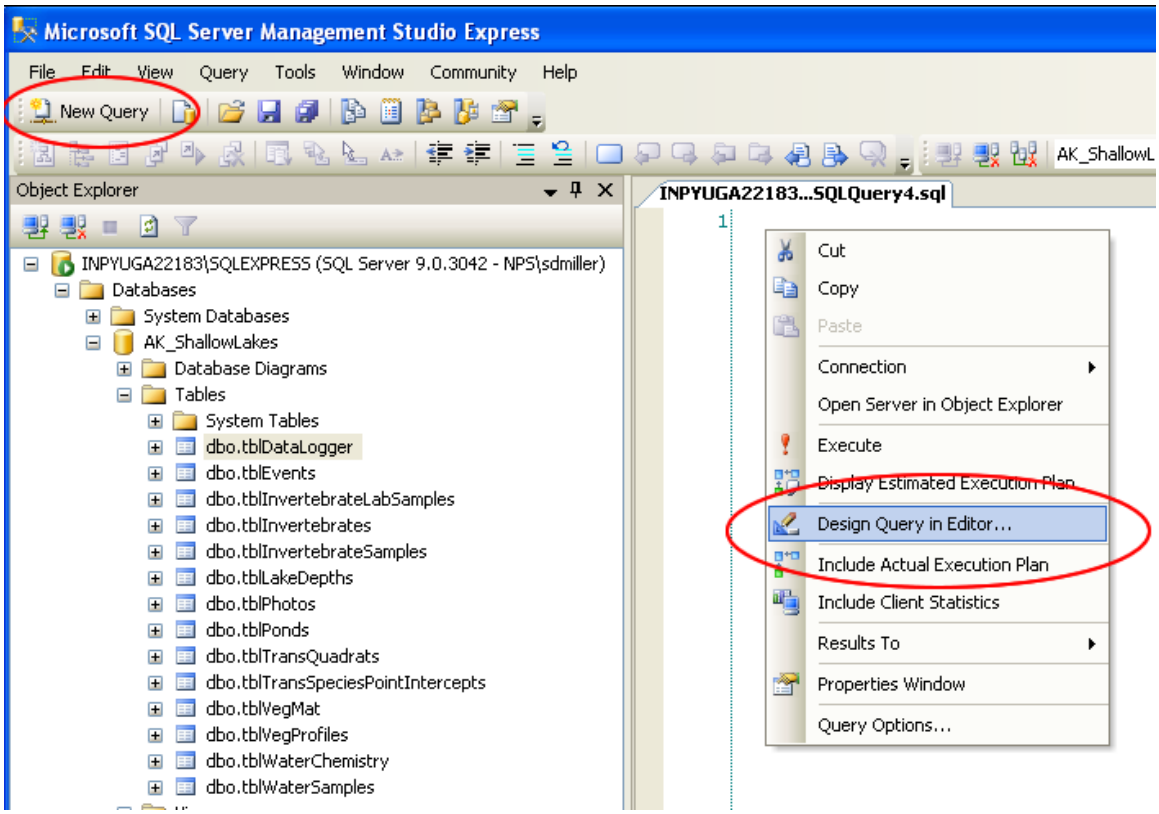
If, in the course of exploratory data analysis, you decide that you have developed a view that you would like to use again on a regular basis, then save it.

**NOTE**: *A view you create in Microsoft SQL Server Management Studio will cascade out to any other front end applications that are linked to the database (i.e., will be available in Microsoft Access Applications as well. Below is a screenshot of a same view seen above in Microsoft SQL Server Management Studio but embedded in a Microsoft Access application. Views are available under the 'Queries' category and may be queried or used in any way you would use a data table. They are a great way to present complex data in a simplified way.*

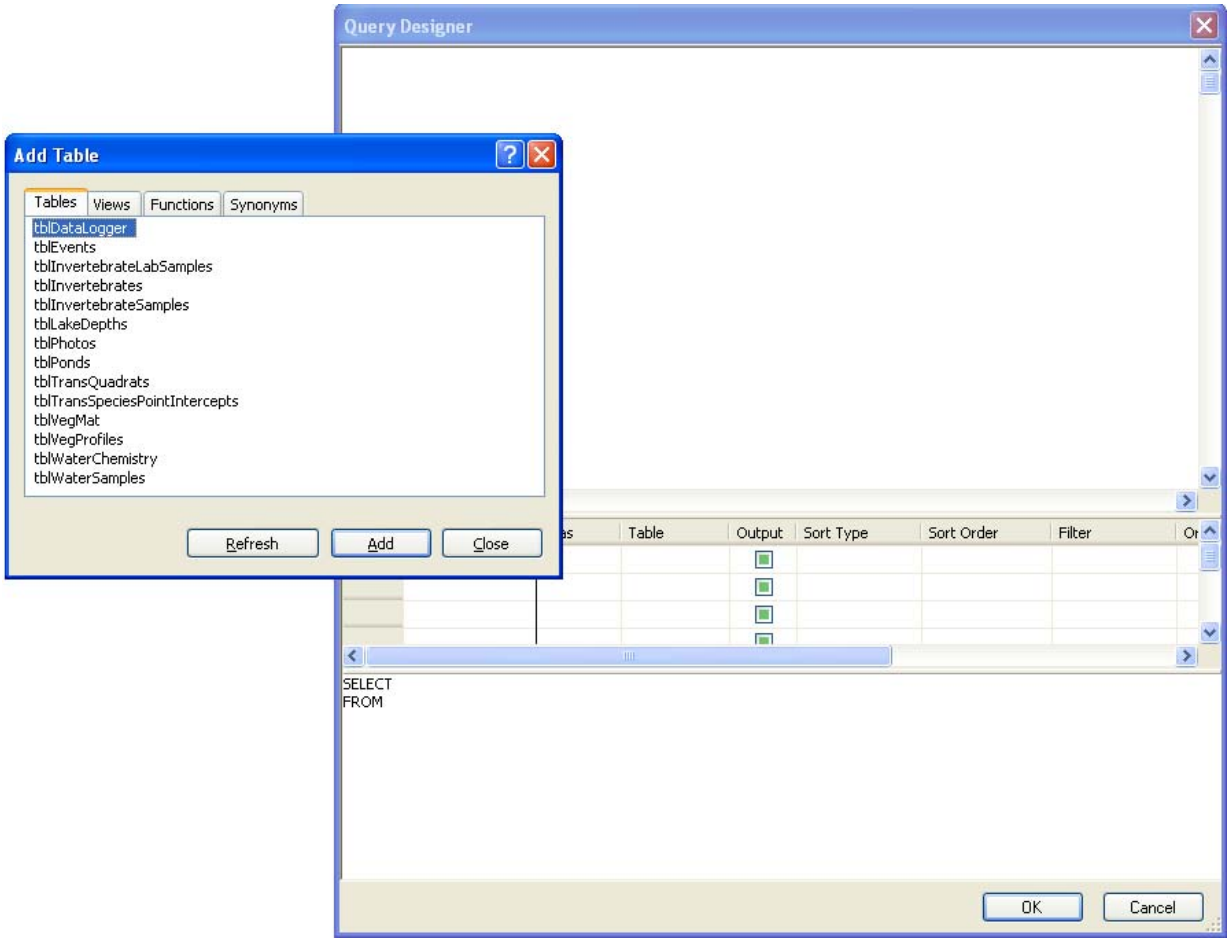



## *How to create a query*

Highlight a table in your database by clicking on it
Select 'New Query' from the toolbar
A tab will appear in the main area to the right
Right click in the tab and select 'Design Query in Editor'

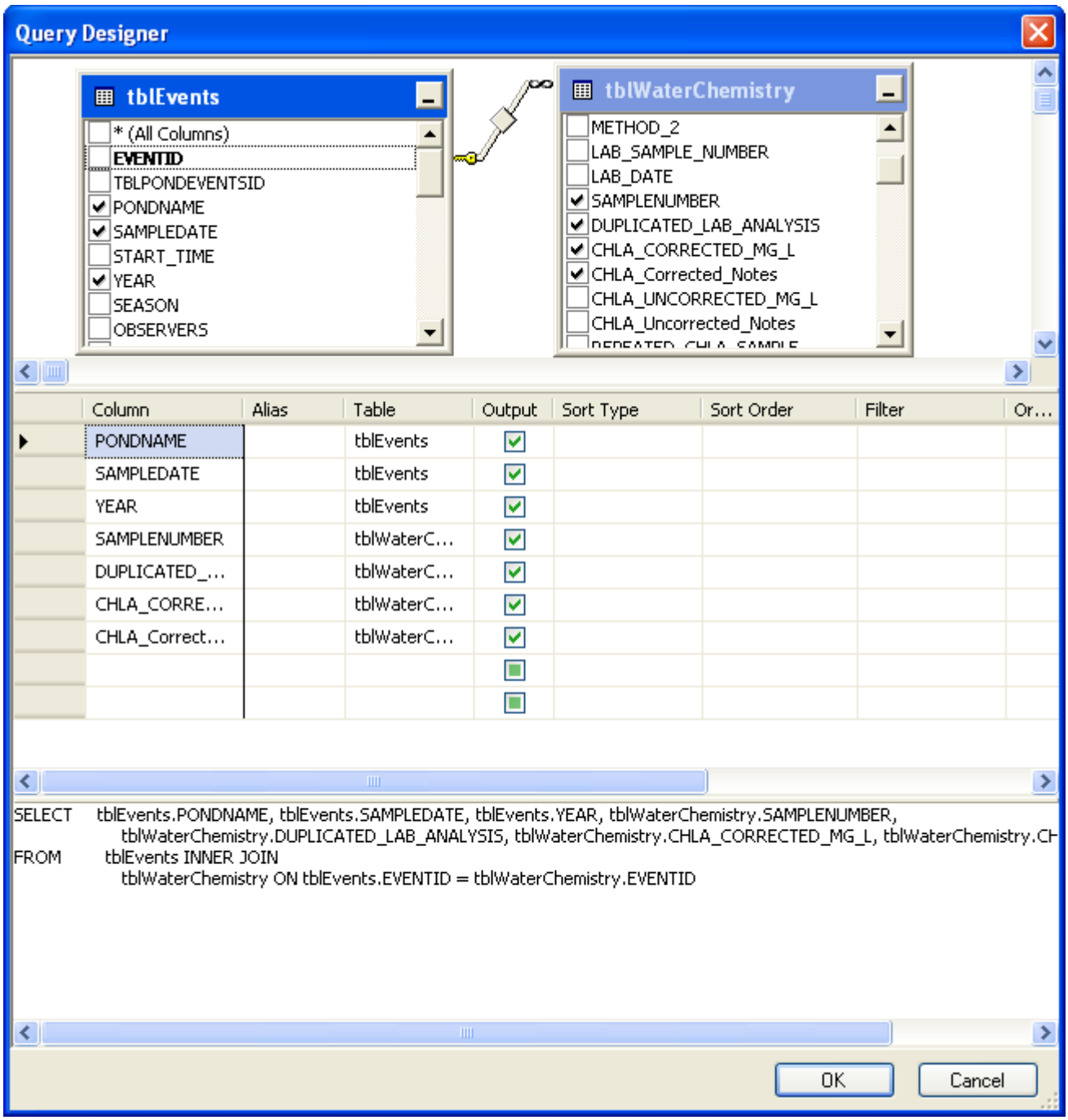



The query designer appears

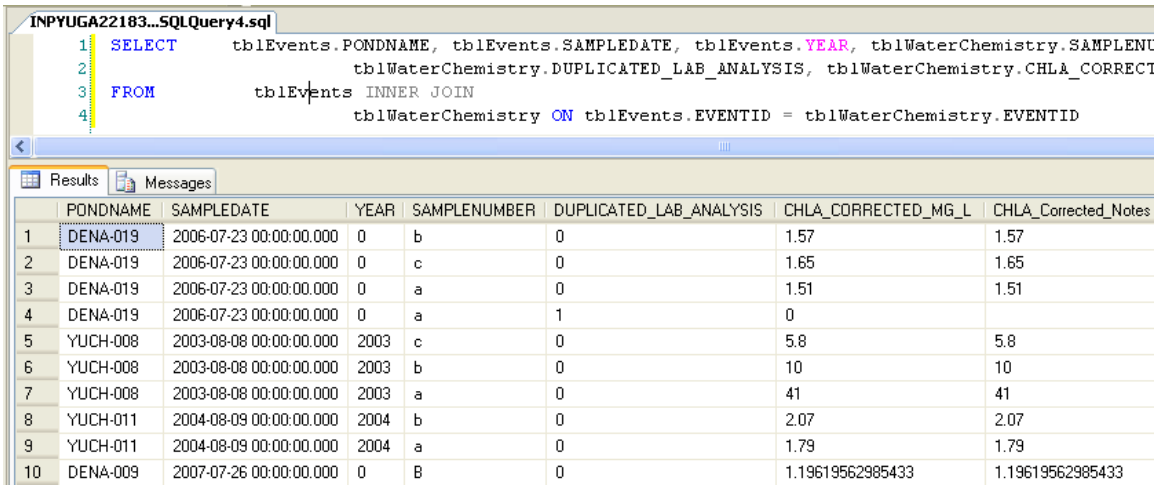



The query designer allows you to select the tables you want to query. The interface is very similar to what you would see in Microsoft Access.

Select your tables.
Use the query designer to build the query and press OK




The query is loaded into the SQL tab.
Click ‘Execute’
The data appears in the grid below.



```
INPYUGA22183...SQLQuery4.sql
1  SELECT     tblEvents.PONDNAME, tblEvents.SAMPLEDATE, tblEvents.YEAR, tblWaterChemistry.SAMPLENU
2                tblWaterChemistry.DUPLICATED_LAB_ANALYSIS, tblWaterChemistry.CHLA_CORRECT
3  FROM          tblEvents INNER JOIN
4                tblWaterChemistry ON tblEvents.EVENTID = tblWaterChemistry.EVENTID
```

Results | Messages

| | PONDNAME | SAMPLEDATE | YEAR | SAMPLENUMBER | DUPLICATED_LAB_ANALYSIS | CHLA_CORRECTED_MG_L | CHLA_Corrected_Notes |
|---|---|---|---|---|---|---|---|
| 1 | DENA-019 | 2006-07-23 00:00:00.000 | 0 | b | 0 | 1.57 | 1.57 |
| 2 | DENA-019 | 2006-07-23 00:00:00.000 | 0 | c | 0 | 1.65 | 1.65 |
| 3 | DENA-019 | 2006-07-23 00:00:00.000 | 0 | a | 0 | 1.51 | 1.51 |
| 4 | DENA-019 | 2006-07-23 00:00:00.000 | 0 | a | 1 | 0 | |
| 5 | YUCH-008 | 2003-08-08 00:00:00.000 | 2003 | c | 0 | 5.8 | 5.8 |
| 6 | YUCH-008 | 2003-08-08 00:00:00.000 | 2003 | b | 0 | 10 | 10 |
| 7 | YUCH-008 | 2003-08-08 00:00:00.000 | 2003 | a | 0 | 41 | 41 |
| 8 | YUCH-011 | 2004-08-09 00:00:00.000 | 2004 | b | 0 | 2.07 | 2.07 |
| 9 | YUCH-011 | 2004-08-09 00:00:00.000 | 2004 | a | 0 | 1.79 | 1.79 |
| 10 | DENA-009 | 2007-07-26 00:00:00.000 | 0 | B | 0 | 1.19619562985433 | 1.19619562985433 |

### *Saving a query*

Queries are saved as text files with .sql extension. Select File → Save As…

# References

# About This Standard Operating Procedure

**Version**: 1.0
**Status**: Draft
**Publication Date**: February 11, 2009
**Author(s)**: Scott D. Miller, Data Manager, Arctic Network Inventory & Monitoring Program.
**Abstract**: Arctic Network uses SQL Server as an enterprise database for storage and retrieval of monitoring data. Where analysis is routine, many queries and analyses can be built into the SQL Server or any of a variety of front end applications. Exploratory data
**Suggested Citation**: NPS-ARCN (2007).Exploratory Data Analysis Using SQL Server Management Studio Version 1.0, Arctic Network-Inventory and Monitoring Program, National Park Service. Fairbanks, Alaska.

# Revision History

| Version | Version Date | Revised By | Changes |
|---|---|---|---|
| 1.0 | 20090129 | S. Miller | Original |
| | | | |

This table reflects changes to this document. Version numbers will be incremented by one (e.g., Version 1.3 to Version 2.0) each time there is a significant change in the

process and/or changes are made that affect the interpretation of the data. Version numbers will be incremented after the decimal (e.g., Version 1.6 to Version 1.7…1.10….1.21) when there are changes to grammar, spelling, or formatting, or minor modifications in the process that do not affect the interpretation of data.